For reprint orders, please contact: reprints@futuremedicine.com

Chromosome X genomic and epigenomic aberrations and clinical implications in breast cancer by base resolution profiling

Aim: Abnormal inactivation or loss of inactivated X chromosome (Xi) is implicated in women's cancer. However, the underlying mechanisms and clinical relevance are little known. **Materials & methods**: High-throughput sequencing was conducted on breast cancer cell lines for copy number, RNA expression and 5'-methylcytosine in ChrX. The results were examined in primary breast tumors. **Results & conclusion**: Breast cancer cells demonstrated reduced or total loss of hemimethylation. Most cell lines lost part or one of X chromosomes. Cell lines without ChrX loss were more active in gene expression. DNA methylation was corroborated with Xi control lincRNA XIST. Similar transcriptome and DNA methylation changes were observed in primary breast cancer datasets with clinical phenotype associations. Dramatic genomic and epigenomic changes in ChrX may be used for potential diagnostic or prognostic markers in breast cancer.

Keywords: breast cancer • Chromosome X • differential gene expression • differential methylation • DNA methylation • DNA sequencing • patient survival • RNA sequencing • RRBS

In normal female cells, one of the X chromosomes is deactivated to counter the potential double dose effect of gene expression as male cells only have one Chromosome X. The inactivation is initiated by the transcription of the long noncoding XIST RNA on the to-be-inactivated Chromosome X. The XIST spreads on the chromosome and the coating provides the template for a series of histone modifications such as histone methylation and deacetylation. This process is coordinated by the X-chromosome inactivation center (XIC). The end result is that about 85% of the genes on the inactivated Chromosome X (Xi) are silenced and remaining 15% escape the inactivation [1,2].

Abnormal inactivation or loss of Xi is implicated in X-linked diseases and women's cancer [1,2]. Skewed Chromosome X inactivation has been implicated in *BRCA1* mutations [3] and breast cancer [4,5]. It has been reported that the Barr body (condensed Xi) may be lost in breast cancer. Tumors lacking Xi may also present a duplication of the active Chromosome X (Xa) or reactivation of Xi. These observations are mostly obtained from low resolution technologies such as FISH or aCGH and it is not clear how these changes are reflected at molecular base level such as DNA methylation and gene expression and what are their biological implications.

To gain a better understanding of Chromosome X abnormalities, we profiled eight breast cell lines (one from normal and seven from cancer) through base resolution next-generation sequencing for their DNA copy number, RNA expression and 5'-methylcytosine modification of DNA. The characteristic changes from the cell lines were further investigated in two sets of breast cancer tissue samples. Through the integrative analysis, we found dramatic DNA copy, DNA methylation and gene expression changes in Chromosome X of breast cancers. Genes that were changed in both cell lines and solid tumors were identified and their clinical implications were explored.



Epigenomi



kocher.jeanpierre@mayo.edu



Materials & methods

Breast cancer & normal epithelial cell lines

Seven breast cancer cell lines (BT474, MCF7, T47D, ZR75, BT20, MDAMB231, MDAMB468) and one nontumor breast epithelium cell line (MCF10A) were sequenced for mRNA (mRNA-seq), DNA (DNAseq) and reduced representation bisulfite sequencing (RRBS) by Illumina Genome Analyzer IIx sequencer. Among the seven cancer cell lines, four are estrogen receptor positive (BT474, MCF7, T47D, ZR75) and three are negative (BT20, MDAMB231, MDAMB468). The former were also characterized as luminal and the latter as basal subtype according to gene expression profiling through microarray [6]. The phenotypic information for the cell lines is summarized in Supplementary Table 1 (for supplementary information please see online at: www.futuremedicine. com/doi/full/10.2217/epi.15.43).

Sequence data analysis of the eight cell lines

The detailed library construction, sequencing and data preprocessing for DNA-seq, mRNA-seq and RRBS were described previously [7]. Briefly, the sequence reads were mapped to University of California Santa Cruz (UCSC) human reference genome (hg19) to obtain normalized gene expression counts, reads per kilobase per million mapped reads (RPKM), for mRNA-seq, DNA CNAs with reference to MCF10A from low coverage DNA-seq, and methylation status of 1,023,020 CpG sites that were commonly captured in the eight cell lines with $\geq 10 \times$ coverage from RRBS. The validation of the RNA-seq data was conducted with NanoString nCounter platform and RT-PCR for a subset of genes and high concordance was obtained [7]. The analyses were focused on the Chromosome X, which includes 1056 annotated genes (including miRNAs) and 16,869 CpG sites. Differentially expressed genes and methylated CpGs were identified by comparing each cancer cell line to MCF10A to get the log2 fold change of RPKM and methylation ratio difference and subsequently through one sample T statistics (assuming no difference between cancer and normal cell line). The genes or CpGs with p-value less than 0.05 were suggestive of differentially expressed or methylated. According to a CpG location relative to nearby genes, CpGs were classified into three categories: within 5 kb of a gene transcription start-site (TSS), greater than 5 kb downstream of a TSS but still within gene body, or otherwise with no nearby gene structures (intergenic regions). Only the first two categories of CpGs were used in the integrative analysis. The raw sequence data are accessible through GEO with accession number GSE27003.

DNA copy number by aCGH

The array CGH experiment was carried out previously for the eight breast cell lines [6] and the CNAs identified from that study were first investigated; however, in that study, no significant CNA abnormalities were reported for Chromosome X in any of the cell lines and we suspected these were likely the result of stringent CNA call parameters applied (cghFLasso with FDR <1%). We then downloaded the log2 ratio data for the cell lines and conducted CNA analysis by genomic segmentation algorithm implemented in the Partek Genomics Suite ([8], MO, USA) with the minimum number of ten markers, p-value threshold of 0.001 and gain or loss more than a half copy.

Breast cancer data from public database

We downloaded two public datasets. The first consists of 16 breast tumor and normal tissues, 8 from each, profiled for DNA methylation with Illumina 450 k microarray (GEO GSE29290). The data were normalized for probe I and II bias [9]. The second dataset was the level III Illumina methylation 450 k and mRNA-seq data for 74 patients with both tumor and normal tissues for both DNA methylation and mRNA-seq data from The Cancer Genome Atlas (TCGA) [10]. The raw data were processed by 'methylumi' or 'minfi' R packages where methylated (M) and unmethylated (U) summary intensities were background adjusted with normalexponential deconvolution. Multiple-batch archives had the intensities in each of the two channels multiplicatively scaled to match a reference sample. The derived summary measure, i.e., beta values from M/(M+U) for each interrogated locus was annotated for UCSC hg19. Probes having a SNP within 10 bp of the interrogated CpG site or having 15 bp from the interrogated CpG site overlap with a REPEAT element were masked as NA. These CpGs along with the ones having a nondetection probability greater than 0.05 in any sample were excluded for further analysis. A total of 386,520 CpGs in all chromosomes ended up in the final analyses after various filtering steps. The final dataset for Chromosome X had 9653 CpGs. The differentially methylated CpGs (DMCs) were detected using paired t statistics. The level III normalized gene level expression (mRNAseq data v2) was estimated from RSEM algorithm [11] and normalized to the upper quartile count at 1000. Differentially expressed genes were identified using R package limma model [12] between tumor and normal pairs. Genes associated with overall survival of patients were identified using Cox proportional hazard model.

Results

Loss of hemimethylation in Chromosome X is characteristic of breast cancer cells

Among the eight cell lines sequenced, there were 1,023,020 common CpGs with at least 10× coverage, of which 16,869 were mapped to Chromosome X. There were 13,140 CpGs located within 5 kb of the transcription start site (TSS) or gene body of 575 genes and 3729 CpGs outside of the genic regions. In normal XX genotype cells, a significant proportion of CpGs is semimethylated in Chromosome X as one of them goes through Xi as the result of noncoding RNA wrapping, DNA methylation and other epigenetic modifications. This indeed was the case for the normal breast cell line MCF10A (Figure 1A, red line) where a trimodal distribution was observed with the highest peak centered around 40%. However, this trimodal distribution was totally or partially lost in all cancer cell lines where the central peak shifted or merged with the left unmethylated peak (a bimodal pattern, Figure 1A). This distribution shift reflected the dramatic methylation reduction in cancer cell lines. Although it is common that cancers demonstrate global hypomethylation and localized hypermethylation in the promoter region, the dramatic shift is rarely seen in autosome chromosomes and methylation distributions in autosome chromosomes were indistinguishable in cancer and normal cell lines (Figure 1B for Ch1). To explore where the CpG methylation changes occurred, we plotted the summarized methylation using local regression for all CpGs within upstream 5 and 15 kb downstream for each sample. Over 77% of CpGs were located within 5 kb of TSS (gray density plot in Figure 1C). Compared with normal cell line MCF10A, cell lines BT20, BT474, MDAMB231, MDAMB468 and T47D had little overall methylation from 5 kb upstream to 1 kb downstream, whereas MCF7 and ZR751 had significantly reduced methylation around TSS compared with MCF10A. For the cancer cell lines, the methylation in gene body appeared to be increased relative to the normal cell line (Figure 1C).

We conducted differential methylation analysis for individual CpGs in Chromosome X using one sample *t* test. There were 11,856 CpGs within 5 k of TSS or within gene body, of which 9373 CpGs (79% of total) were differentially methylated in cancer cell lines compared with the normal cell lines at p-value < 0.05 and 6790 were with methylation difference >10% additionally (57%). These DMCs (6790 with p < 0.05 and mean difference >10%) were mostly concentrated around 1 kb of TSS (5243, 77%) with involvement of 561 genes; the vast majority of these CpGs were hypomethylated (6012 out of 6790, 89%, Figure 1D). For most cell lines, the reduced methylation mainly occurred in the CpGs that used to be hemimethylated (Figure 1E & F for MCF10A and BT20 where the middle band of DNA methylation is dramatically reduced compared with MCF10A).

DNA copy change or loss/reactivation of Xi in the breast cancer cells

We conducted copy number aberration (CNA) analysis for the seven cancer cell lines using the noncancer cell line MCF10A as a baseline for the whole genome DNA sequencing data [7]. These data showed that cell line BT20 had long intermittent deletion in the p arm and continuous deletion in half of the q arm close to the centromere, BT474 had sporadic deletion in the q arm, T47D has whole q arm loss and MDAMB231 had deletion in the majority of the q arm and part of the p arm (Supplementary Figure 1). To cross validate the result, we re-analyzed the aCGH data [6] for the same seven cancer cell lines using more lenient criteria (no CNAs were reported in the original report using FDR cutoff < 0.01), and four deletions were identified in Chromosome X of cell line BT20, BT474 and T47D. These observations were consistent with the DNA-seq CNA data, except that the cell line MDAMB231 had no deletion detected in the aCGH data. A few localized amplifications were observed in BT474, MDAMB468 and ZR751 in the sequencing data but not in the aCGH data (Supplementary Figure 1).

The dramatic methylation change pattern in Chromosome X could be explained: first, Xi loss or deletion. The Xi is highly condensed through lncRNA wrapping, DNA methylation and other mechanisms in the normal cells. The loss of Xi would bias the results in favor of the Xa which is largely unmethylated and active; and second, reactivation of Xi or dysfunction of Xi process would also lead to the similar changes. In this case, the Xi becomes unmethylated and active, which would cause genes expressed higher than when single-active Chromosome X was present. We examined these hypotheses: DNA copy number change; the expression of XIC genes and differentially expressed genes in each cancer cell line compared with the normal cell line.

As presented in the CNA analysis, cell line BT20, BT474, T47D and MDA231 showed partial Chromosome X loss and these cell lines had no or little expression of XIC genes (*XIST, TSIX, JPX* and *FTX*, **Figure 2**) from the RNA-seq data, whereas their overall methylation dramatically shifted to unmethylated state (**Figure 1A**), supporting the conclusion that these cell lines had Xi loss. On the other hand, no major CNAs were observed in cell line MCF7, MDAMB468 and ZR751. The XIC gene expression of these cell lines varied from total loss (MDAMB468) and reduced (MCF7) to no change or increased (ZR751) compared with normal cell line MCF10A (Figure 2), suggesting the Xi partial or total reactivation.

Genes in Chromosome X are more active & upregulated in breast cancer cells

There are 1056 annotated (999 unique) genes (including miRNAs) on Chromosome X. We first

conducted unsupervised clustering and the cancer cell lines were mostly clustered by estrogen receptor (ER) expression status. The noncancer cell line MCF10A had no expression of ER; however, its gene expression pattern in Chromosome X was more like ER-positive cell lines (Figure 3A). The cluster patterns were similar to what was reported previously when genes from all chromosomes were used [7].



Figure 1. Methylation profiles of eight breast cell lines (facing page). (A) Density plot of 16,869 CpGs in Chromosome X. MCF10A (red) is normal cell line with trimodal distribution and most of CpGs are methylated around 40–50%. All others are cancer cell lines and their middle peak is shifted to the left or totally gone. (B) Density plot of CpGs in Chromosome 1. Most CpGs are not methylated and there is no much difference between normal MCF10A (red) and cancerous cell lines (all others), clear contrast with Chromosome X. **(C)** Fitted methylation line (lowess curve) for all CpGs (and genes) around transcription start site (TSS, 0 on X axis) for each cell line. Reduced methylation is most at upstream of genes up to 5 k and within 1 k of downstream. The gray line is the density of CpGs based on their locations. Most are around TSS. **(D)** Differentially methylated CpGs in Chromosome X. The three numbers separated by '/' on the left and right side of figure are the number of CpGs in hyper (left) and hypo (right) methylation in cancer cells. Number of CpGs with p-value <0.05 and mean difference >10%; p-value < 0.05 only; and all CpGs in hyper (>0) or hypo (<0) direction without statistic testing, respectively. **(E)** Chromosome X CpG methylation of normal cell line MCF10A by their genomic locations. Many CpGs are hemimethylated or with range of methylation. **(F)** Chromosome X CpG methylation of cancer cell line BT20 by their genomic locations. Ths: Transcription start-site.

Next, we were interested in identifying consistently changed genes in the cancer cell lines relative to the normal cell line, which would be the cancer genes common to different subtypes of parental cancer cells. To do this, we used the normal cell line as a reference to get the expression differences in log2 scale. One sample *t* test was then applied to test the hypothesis of no significant difference between the difference mean of cancer cell lines from zero. This analysis revealed that 319 genes were significantly changed at the p-value <0.05, 218 up- and 101 downregulated, which were far more than expected if there were no differences (Figure 3B, upper panel). Among these, 86 (up) and 44 (down) were greater than twofold (Figure 3B, lower panel), respectively. The top-up-expressed genes include ALEX family of proteins (ARMCX1, 2, 3, 6), BEX2 and BEX4, AR and POLA1. The top-down-expressed genes generally had a larger fold change and smaller p-value, which include BGN, MSN, HEPH, FHL1, XIST

and TSIX. XIST and TSIX are the genes involved in Chromosome X inactivation and the significant down expression suggests reduced or loss of the function in most of the cancer cell lines. When comparing the differentially expressed genes for each cancer cell line, it is interesting to see that the cell lines without Chromosome X loss but possible Xi reactivation (MCF7, MDAMB468 and ZR751) had the highest up-/down-expressed gene ratios (Figure 3C). Additionally, when overall gene expression was plotted for all genes on the Chromosome X over their genomic coding location, all cancer cell lines had higher expression than the normal MCF10A across the chromosome except the p arm and middle part of q arm in BT20, the q arm of BT474 and MDAMB231 where chromosome deletion was detected from the copy number analysis (Figure 3D). There were 258 genes not detectable in the normal cell line MCF10A but were expressed in one or more cancer cell lines, 27 of which were significantly different from the normal



Figure 2. Relationship of overall DNA methylation and the X-chromosome inactivation center gene (XIST and TSIX) expression in Chromosome X. Overall DNA methylation change in Chromosome X is correlated with the X-chromosome inactivation center noncoding RNA expression of *XIST* and *TSIX*. Normal cell (MCF10A) and some cancer cell lines (MCF7 and ZR751) have high expression of *XIST* and *TSIX* along with high DNA methylation, while other cells have total loss of *XIST* and *TSIX* with significantly reduced DNA methylation. The boxplot is for DNA methylation (the scale on left axis) and the lines are for gene expression (the scale on right axis).



Figure 3. Gene expression profiles and differential expression in cell lines. (A) Unsupervised clustering using all genes in Chromosome X. Samples (cell lines) clustered mostly by their estrogen receptor status. **(B)** Upper panel: p-value distribution of one sample *t* test comparing seven cancer cell lines to the normal cell line. 319 genes with p-value <0.05. Lower panel: Volcano plot of differentially expressed genes from one sample *t* test. Genes with p-value <0.05 and fold change >2 are marked as red. **(C)** Differentially expressed genes from each cell line compared with the normal cells. Fold change >2 was used as the cutoff. The number in the box is the number of genes meeting the cutoff. Orange box for genes expressed higher; green for genes expressed lower and white for genes without significant change. **(D)** Fitted gene expression trend across Chromosome X for eight cell lines. Most cell lines had higher gene expression across Chromosome X than the normal cell line (MCA10A). FC: Fold change.

cell line with p-value <0.05. These would be the genes that were activated in the cancer. Most of these genes were expressed at very low level and seven had increased expression greater than twofold, including ARMCX1, 2, 6, BEX2, 4, GYG2 and GSF2RA (Supplementary Table S2). BEX2 is a regulator of mitochondrial apoptosis and G1 cell cycle [13]. Evidence has shown that this gene is implicated in breast cancer apoptosis. Increased expression of this protein may protect the breast cancer cells against apoptotic process for uncontrolled growth [13]. BEX2 appears also promoting cell migration and invasion of glioma cells [14]. The 5'-UTR or promoter of this gene had significantly reduced methylation in the cancer cell lines, supporting that the gene reactivation may be the result of demethylation in DNA.

Methylation pattern changes in solid breast cancer & normal tissues (GSE29290)

To access whether the characteristic DNA methylation changes of Chromosome X in breast cancer cell lines are present in primary breast tumors, we downloaded a methylation dataset from GEO with accession number GSE29290 [15]. The dataset has eight breast tumor and eight normal samples profiled with the Illumina Human Methylation 450 k platform, which contains over 480-k-targeted CpGs across the genome. The data were normalized to correct infinium I and II chemistry bias using peak adjustment [9]. We extracted 11,230 CpG sites on Chromosome X and excluded 16 CpG sites with incomplete beta values across the 16 samples, which led to 11,185 CpGs in the final analyses. As shown in the density plot of Figure 4A, the CpG methylation in the



Figure 4. DNA methylation profiles in breast cancer and normal for eight tumor and normal pairs. (A) Density plot of Chromosome X DNA methylation. All tumors have reduced methylation with some more dramatic compared with normal tissues. X-axis for methylation beta value range from 0 to 1. (B) Differentially methylated CpGs between tumors and normal. Much more DMCs (five fold, 2561 vs 533) are hypomethylated in tumors compared with normal tissues. X-axis for methylation beta mean difference between tumor and normal samples.

cancer samples shifted to lower methylation, with the central peak moving away from the expected 0.5, similar to the patterns seen in breast cancer cell lines. However, the patterns were less dramatic and more variable in the tumors, mostly likely because they generally contain variable amounts of normal tissue, which could dilute cancer methylation changes. Additionally, the methylation profile in the tissue samples from Illumina 450 k did not have peaks centered around 0 or 1, which could be the results of: the microarray platform is generally not sensitive to detect methylation at 0 (totally unmethylated) or 1 (totally methylated); while RRBS biases CpGs in CpG islands and promoter regions, the CpGs in the microarray are most selective in the genic or regulatory regions. The platform differences dictate the profiles from RRBS and 450 k microarray may not be exactly the same. In the subsequent differential methylation analysis, we found roughly fivefold more hypomethylated CpGs than hypermethylated CpGs at p-value less than 0.05 (2561 vs 533, Figure 4B), the similar finding as seen in the breast cancer cell lines. About 74% of these DMCs are in CpG islands, shores or shelf. For the 561 genes with DMCs in the cell lines, 498 (89%) were also differentially methylated in these primary tumor tissues.

TCGA paired tumor & normal samples with both methylation & mRNA-seq data

To explore the gene expression and methylation data simultaneously in patient samples, we further analyzed the breast cancer dataset with 74 tumor and normal pairs from TCGA database. We intended to use this data to compare with cell line data for common changes that might be clinically relevant. As seen in the cancer cell lines, many tumors demonstrated the similar reduced DNA methylation and shift in Chromosome X (Figure 5A). Again, these changes were more subtle for some of the tumors and there was more variability from tumor to tumor, where some more closely resembled the changes seen in cancer cell lines and others were more like normal tissues. Interestingly, for one patient with both a primary and a metastatic tumor, the metastatic tumor had much more dramatic reduction of DNA methylation in Chromosome X compared with its primary tumor (Figure 5B), suggesting the metastatic nodule might have derived from a subclone with different genomic abnormalities and metastatic potential than the primary tumor. In the DMC analysis, 6477 out of 9653 (67.1%) CpGs in Chromosome X were significantly different between paired tumor and normal tissues, the highest percentage among all chromosomes (Figure 5C). Among the DMCs, over 65.4% were hypomethylated (4235 vs 2242) in Chromosome X, while all other chromosomes except chromosome 8 were more hypermethylated (Figure 5D). In the analysis of tumors, unsupervised clustering using 2382 CpGs with standard deviation greater than 0.4 across samples classified the 74 samples into three major groups and the grouping was significantly associated with tumor progesterone receptor (PR) positivity status (Fisher's p-value < 0.01) and borderline associated with tumor TNM stage (Fisher's p-value = 0.09).

In the RNA-seq data analysis, unsupervised clustering for all tumors and normal samples showed clear separation between tumor and normal tissues. When unsupervised clustering was applied to tumors only, there were two major subclusters, one mainly consisting of ER positive (Fisher's p-value < 0.01), PR positive (Fisher's p-value < 0.01) and luminal A type (PAM50 classification, Fisher's p-value < 0.01) with more patient alive and another with mostly basal cell type and more patients deceased. The clustering appeared not significantly associated with the tumor TNM stage (Fisher's exact p-value 0.87). Differential expression analysis between tumors and normal samples



Figure 5. DNA methylation profiles of 74 pairs of tumor and normal from The Cancer Genome Atlas. (A) Density plot of Chromosome X DNA methylation for 74 pairs of breast tumor and normal samples. Most tumors have reduced methylation. The methylation is much more variable in tumors. **(B)** Methylation profiles of a primary tumor, its metastasis and adjacent normal breast, showing increasing reduced DNA methylation. **(C)** Proportion DMCs per chromosome where Chromosome X had the highest proportion of DMCs. **(D)** All chromosomes but 8 and X had more hypermethylated than hypomethylated CpGs. Chromosome X had much more hypomethylated CpGs.Y-axis for number of CpG sites in hyper (above 0 line) or hypo (below zero line) direction. DMCs: Differentially methylated CpGs.

revealed 491 genes with FDR less than 0.05 (61.6% genes), of which 217 were higher and 275 were lower in the tumors than in the normal breast tissues. Like breast cancer cell lines, the primary tumors had significantly reduced *XIST* expression (more than twofold), while *JPX* and *TSIX* had little change (Figure 6A). As observed in cell lines, the *XIST* expression is positively correlated with Chromosome X overall (median) methylation (Pearson correlation coefficient 0.27 with p-value = 0.02, Figure 6B).

Commonly changed genes between cell lines & solid tumors

Cancer cell lines *in vitro* may have gene expression changes unique to culture conditions but not relevant

to tumors *in vivo*; on the other hand, solid tumors may have changes as the result of different stromal or infiltrating tissue contamination. We reasoned that genes commonly changed in both may represent specific gene changes in breast cancer. We compared the genes significantly changed in the cell lines and in the primary tumor samples and found 212 commonly changed genes in Chromosome X, among which 162 were (76.4%) changed in the same direction (Figure 7, Supplementary Table 3). Pathway analysis for the 162 genes showed the top enriched pathways include metabolic biosynthesis (PRPP, cholesterol biogenesis) and mitochondrial dysfunction and the top involved disease and networks include hereditary, developmental and neurological disorders. Notably, the genes that are nor-



Figure 6. The X-chromosome inactivation center gene expression and its association with DNA methylation in Chromosome X. (A) X/ST is significantly down expressed in tumors (reduced 2.23-fold, p-value 4.35E-09), while JPX and TS/X do not have much difference between tumors and normal samples. RPKM Reads per kilobase per million mapped reads. (B) X/ST expression is positively correlated with Chromosome X median methylation.

mally repressed but reactivated in cancer cell lines such as *ARMCX* and *BEX* family were among the discrepancy genes. For example, *ARMCX1*, 2 and 6 were highly expressed in cancer cell lines, but they were not differentially expressed (*ARMCX 2* and 6) or even lower in tumor tissues. The same were true for *BEX2* and *BEX4*. These data illustrate the complexities of using model systems and tissue mixture of clinical tumor samples.

Commonly changed genes & clinical outcomes

To look into if the cancer-specific genes were also associated with clinical outcome of patient survival time, we conducted Cox proportional hazards model analysis for the commonly differentially expressed genes both in the cell lines and the tumor samples from TCGA. Among the 212 genes, 21 were significantly associated with survival at p-value less than 0.05 (Table 1), some of which were reported previously for their clinical relevance to breast cancer patient outcomes. For example, MOSPD1 plays an important role in epithelial to mesenchymal switch [16], an indication of more aggressive tumor behavior; AMOT promotes the proliferation of mammary epithelial cells [17]; Overexpression of PDK3 increases drug resistance and early recurrence in colon cancer [18]; and RLIM is an important ERa cofactor and is correlated with clinical phenotypes of breast cancer [19]. These genes were highly expressed in tumors with significantly reduced CpG methylation in their promoter region. On the protective side, HDAC6 higher expression is found to be associated with small, low grade, estrogen-positive tumors and better survival in breast cancer [20] and RBM10 collaborates with proapoptotic BAX gene and Caspase-3 to inhibit tumor growth of breast cancer [21,22].

Discussion

Chromosome X abnormality has been observed in various genetic disorders and cancers. However, its mechanism, roles in cancer development and clinical implications are little known. In this study, we investigated the DNA methylation, gene expression and copy number abnormalities of breast cancer cell lines and then extended to primary breast tumors. Widespread genomic changes in cancer cells were observed, which include dramatic DNA methylation shift, partial or total loss of Xi, reactivation of Xi and overexpression of majority of the genes in Chromosome X. Similar



Figure 7. Commonly changed genes between cell lines and primary tumors. For the genes that are differentially expressed commonly between cell lines and primary tumors, majority of them (76.4%) are in the same changed directions and these genes are more likely tumor specific.

TGCA: The Cancer Genome Atlas.

Table 1. Differentially expressed genes in both cell lines and primary tumors in The Cancer Genome Atlas and their survival association of expression in tumors.						
Gene	Log2 FC cell line	P _{val} cell line	Log2 FC tissues	P _{val} tissues	HR in tumors	P _{val} in tumors
MOSPD1	0.82	0.02616	0.32	3.06E-06	4.78	0.0023
PHF6	0.80	0.02298	0.37	9.36E-06	3.72	0.0075
RLIM	0.65	0.00519	0.20	0.001148	2.56	0.0397
RAP2C	1.03	0.04225	0.93	1.07E-14	2.50	0.0107
ZDHHC9	0.88	0.02537	0.66	2.29E-14	2.42	0.0147
PDK3	1.44	0.00187	0.63	5.78E-07	2.39	0.0135
GPC4	-1.34	0.03921	0.76	6.74E-06	2.19	0.0002
AMOT	1.90	0.03266	-0.66	5.83E-06	2.16	0.001
WBP5	-2.30	0.02012	-0.33	0.001225	2.10	0.0287
HEPH	-4.24	6.73E-14	-0.47	0.001181	1.71	0.012
SRPX	-2.42	0.0089	-2.61	4.92E-25	1.55	0.0427
TMEM47	-2.22	2.27E-05	-1.64	3.94E-18	1.36	0.0372
PHEX	1.26	0.01265	0.89	0.000177	1.30	0.0376
CLDN2	0.39	0.03424	-2.17	3.25E-12	1.26	0.0214
CXCR3	-0.08	0.00704	0.81	8.12E-05	0.82	0.041
FAM50A	1.11	4.98E-05	0.56	8.54E-07	0.49	0.02
GPKOW	0.75	0.00657	0.24	0.001015	0.49	0.0299
GRIPAP1	0.71	0.00598	0.32	0.000441	0.47	0.0046
RBM10	0.68	0.00194	0.25	0.000446	0.44	0.0305
HDAC6	0.83	0.00211	-0.16	0.002709	0.32	0.0284
CCDC22	1.20	0.0007	0.41	3.74E-08	0.31	0.0029

The genes are sorted by descending order of hazards ratio (HR). HR above 1 indicates higher expression of a gene correlates with worse patient survival and below 1 for better survival.

FC: Fold change; HR: Hazards ratio; P_{val}: P-value.

changes were seen in breast tumors; however, these changes were more subtle and variable and in some cases may not agree, which may not be surprising as tumor tissues almost always contain some or sometimes significant amount of normal components. The normal cell contamination inevitably obscures the tumor-specific changes. In spite of this, many similar genomic changes between cell lines and solid tumors were found from this study.

Chromosome X gain is reported in some cancers. The gain and increased expression of androgen receptor (AR) from Chromosome X in prostate cancer is associated with prostate cancer progression and recurrence [1]. However, in breast cancer, the change appears to be mainly associated with loss, although there may be reactivation of Xi or gain of active X exists [23]. Both cases can lead to globally reduced DNA methylation and the latter may cause gene expression dosage imbalance. A report showed that loss of Chromosome X locus (at Xq25) was seen in 52% of (out of 72) infiltrating ductal carcinomas [24]. This loss was correlated with a larger tumor size, higher his-

tological grade and axillary lymph-node metastasis. In our cell line data, the cell lines with Chromosome X loss or dramatic DNA methylation change mostly had ER negative and more aggressive phenotypes, supporting the observation that the abnormality of Chromosome X not only plays a role in breast cancer development but also tumor progression [2,25].

Cancer cells *in vitro* can have genomic changes as a result of culture status which may not be relevant to tumors. Solid tumors are mostly contaminated by various normal tissue components. By comparing the commonly changed genes, it is more likely we can find the changes that are more tumors specific. Indeed, we have found a set of genes that were commonly differentially expressed and majority of them changed in the same directions. These genes may play an important role in tumor development, some of which may lead to varied tumor phenotypes. For example, the lncRNA *XIST* was reduced or lost in both cell lines and solid tumors and experimental data showed that it is not only required for maintaining Xi but also a potent suppressor of hematologic cancer [26]. The change of XIST was also found to be associated with cancer chemotherapy response [27]. FOXP3 was increased in cancer cell lines (1.5-fold; p-value = 0.02) and in tumor tissues, it was increased more than fourfold (p < 0.0001) in our analyzed datasets, suggesting its important role in breast cancer development although not observed for survival association. All eight CpGs in the upstream of the TSS were hypomethylated in tumors. Interestingly for DNA methylation in cell lines, all CpGs upstream of the gene were significantly demethylated (27–39%) and all CpGs in the gene body were hypermethylated (17-57%). Recent studies have shown that FOXP3 can be expressed in epithelial cells and overexpression of the gene is associated with poor outcome and metastasis [28,29].

The loss or reactivation of Xi leads to the similar DNA methylation changes; however, their impact on gene expression appears different: the reactivation gave rise to more genes expressed or highly expressed. It is possible that the change can be very tumor specific and studies to distinguish the genomic abnormality for each tumor would like to help identify unique changes to a tumor for targeted therapy or individualized care.

Conclusion

Chromosome X in breast cancer has widespread genomic abnormalities. The inactivated Chromosome X can be lost or reactivated which may lead to dramatic DNA methylation and unbalanced gene expression, which could be potentially used for diagnostic testing. These changes likely contribute to carcinogenesis, varied tumor phenotypes or clinical outcomes that warrant further investigation.

Financial & competing interests disclosure

This work was supported by Division of Biomedical Statistics and Informatics Meritorious Award (to Z Sun) and Mayo Clinic Center for Individualized Medicine. Additional support was derived from the 26.2 with Donna Foundation. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The breast cancer datasets used in this study were downloaded from public database GEO and TCGA where the data providers had appropriate institutional review board approval for their studies. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved. The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Open Access

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/bync-nd/3.0/

Executive summary

- One of X chromosomes in normal female cells is inactivated through epigenetics modifications.
- The inactivated X chromosome can be either partial or totally lost in some breast cancers.
- The epigenetics modifications of the inactivated X chromosome are not maintained in other breast cancers.
- The genetic and epigenetic changes of the X chromosome lead to dramatic DNA methylation shift and gene expression changes.
- DNA methylation is dramatically reduced in the X chromosome.
- The X chromosome inactivation center genes are reduced or not expressed.
- The common gene changes between breast cancer cell lines and primary tumors exist that affect tumor behaviors and patient survival.

References

Papers of special note have been highlighted as: • of interest

- Spatz A, Borg C, Feunteun J. X-chromosome genetics and human cancer. *Nat. Rev. Cancer* 4(8), 617–629 (2004).
- Chaligne R, Heard E. X-chromosome inactivation in development and cancer. *FEBS Lett.* 588(15), 2514–2522 (2014).
- 3 Lose F, Duffy DL, Kay GF, Kedda MA, Spurdle AB. Skewed X chromosome inactivation and breast and ovarian cancer status: evidence for X-linked modifiers of BRCA1. *J. Natl Cancer Inst.* 100(21), 1519–1529 (2008).
- 4 Kristiansen M, Langerod A, Knudsen GP, Weber BL, Borresen-Dale AL, Orstavik KH. High frequency of skewed X inactivation in young breast cancer patients. *J. Med. Genet.* 39(1), 30–33 (2002).

- 5 Kristiansen M, Knudsen GP, Maguire P *et al.* High incidence of skewed X chromosome inactivation in young patients with familial non-BRCA1/BRCA2 breast cancer. *J. Med. Genet.* 42(11), 877–880 (2005).
- 6 Kao J, Salari K, Bocanegra M *et al.* Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS ONE* 4(7), e6146 (2009).
- Comprehensive profiling of 52 widely used breast cancer cell lines by aCGH and gene-expression microarray for copy number alterations and gene-expression patterns.
- 7 Sun Z, Asmann YW, Kalari KR *et al.* Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS ONE* 6(2), e17490 (2010).
- Comprehensive profiling of seven breast cancer lines by DNA, RNA and methylation sequencing for genomic and epigenomic differences between estrogen-positive and negative-breast cancers.
- 8 Partek. www.partek.com/
- 9 Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the infinium methylation 450k technology. *Epigenomics* 3(6), 771–784 (2011).
- 10 Cancer Genome. http://cancergenome.nih.gov/
- 11 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323 (2011).
- 12 Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article3 (2004).
- 13 Naderi A, Liu J, Bennett IC. BEX2 regulates mitochondrial apoptosis and G1 cell cycle in breast cancer. *Int. J. Cancer* 126(7), 1596–1610 (2010).
- 14 Zhou X, Xu X, Meng Q *et al.* Bex2 is critical for migration and invasion in malignant glioma cells. *J. Mol. Neurosci.* 50(1), 78–87 (2013).
- 15 NCBI. www.ncbi.nlm.nih.gov/geo/
- 16 Thaler R, Rumpler M, Spitzer S, Klaushofer K, Varga F. Mospd1, a new player in mesenchymal versus epidermal cell differentiation. J. Cell. Physiol. 226(10), 2505–2515 (2011).
- 17 Ranahan WP, Han Z, Smith-Kinnaman W et al. The adaptor protein AMOT promotes the proliferation of mammary epithelial cells via the prolonged activation of the extracellular signal-regulated kinases. *Cancer Res.* 71(6), 2203–2211 (2011).

- 18 Lu CW, Lin SC, Chien CW *et al.* Overexpression of pyruvate dehydrogenase kinase 3 increases drug resistance and early recurrence in colon cancer. *Am. J. Pathol.* 179(3), 1405–1414 (2011).
- 19 Johnsen SA, Gungor C, Prenzel T *et al.* Regulation of estrogen-dependent transcription by the LIM cofactors CLIM and RLIM in breast cancer. *Cancer Res.* 69(1), 128–136 (2009).
- 20 Zhang Z, Yamashita H, Toyama T *et al.* HDAC6 expression is correlated with better survival in breast cancer. *Clin. Cancer Res.* 10(20), 6962–6968 (2004).
- 21 Martinez-Arribas F, Agudo D, Pollan M *et al.* Positive correlation between the expression of X-chromosome RBM genes (*RBMX, RBM3, RBM10*) and the proapoptotic Bax gene in human breast cancer. *J. Cell. Biochem.* 97(6), 1275–1282 (2006).
- 22 Martin-Garabato E, Martinez-Arribas F, Pollan M, Lucas AR, Sanchez J, Schneider J. The small variant of the apoptosis-associated X-chromosome *RBM10* gene is coexpressed with caspase-3 in breast cancer. *Cancer Genomics Proteomics* 5(3–4), 169–173 (2008).
- 23 Wang N, Cedrone E, Skuse GR, Insel R, Dry J. Two identical active X chromosomes in human mammary carcinoma cells. *Cancer Genet. Cytogenet.* 46(2), 271–280 (1990).
- 24 Piao Z, Malkhosyan SR. Frequent loss Xq25 on the inactive X chromosome in primary breast carcinomas is associated with tumor grade and axillary lymph node metastasis. *Genes Chromosomes Cancer* 33(3), 262–269 (2002).
- 25 Richardson AL, Wang ZC, De Nicolo A *et al.* X chromosomal abnormalities in basal-like human breast cancer. *Cancer cell* 9(2), 121–132 (2006).
- One of the first few papers describes X chromosome abnormalities and tumor phenotype association.
- 26 Yildirim E, Kirby JE, Brown DE *et al.* Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* 152(4), 727–742 (2013).
- An interesting paper demonstrating *XIST* is not only an important regulator in X inactivation but also a tumor suppressor.
- 27 Weakley SM, Wang H, Yao Q, Chen C. Expression and function of a large non-coding RNA gene *XIST* in human cancer. *World J. Surg.* 35(8), 1751–1756 (2011).
- 28 Triulzi T, Tagliabue E, Balsari A, Casalini P. FOXP3 expression in tumor cells and implications for cancer progression. J. Cell Physiol. 228(1), 30–35 (2013).
- 29 Douglass S, Ali S, Meeson AP, Browell D, Kirby JA. The role of *FOXP3* in the development and metastatic spread of breast cancer. *Cancer Metastasis Rev.* 31(3–4), 843–854 (2012).